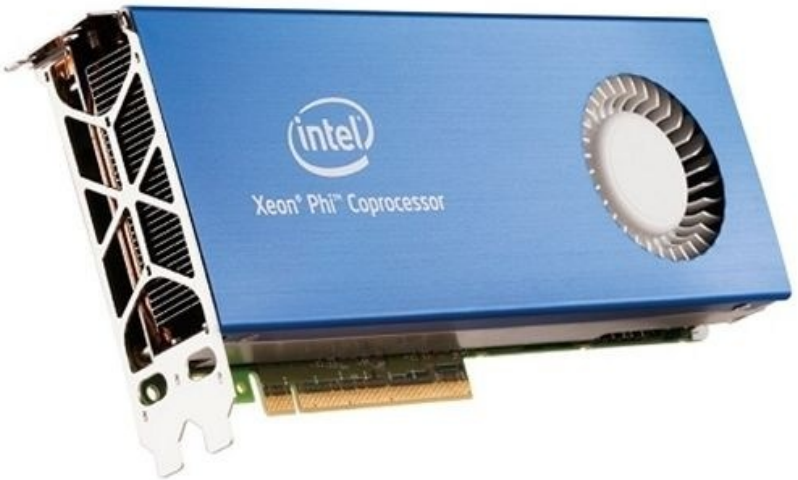
Prototype
Intel Xeon Phi and NVIDIA Kepler Computing Cluster

Don Holmgren
CS/SC/SCF/HPC

All Experimenters Meeting, September 30, 2013

Intel Xeon Phi – Many Integrated Core Architecture (MIC)



Add-in PCIe accelerator card (5110P):

- 60 Pentium-like cores, each with 4 hardware threads
- 8 GB of GDR5 memory with 320 GB/sec peak memory bandwidth
- Cards run Linux; each looks like a 240-core SMP system
- On-board memory is used for both user data, and for a RAM disk for the embedded Linux image
- User programs run “native” on the card, or in “offload mode” (main code runs on the host system, and computational kernels run on the card)
- Most existing C and C++ codes simply need to be re-compiled for MIC. However, performance will be only a fraction of the capability of the card without tuning or modifications.

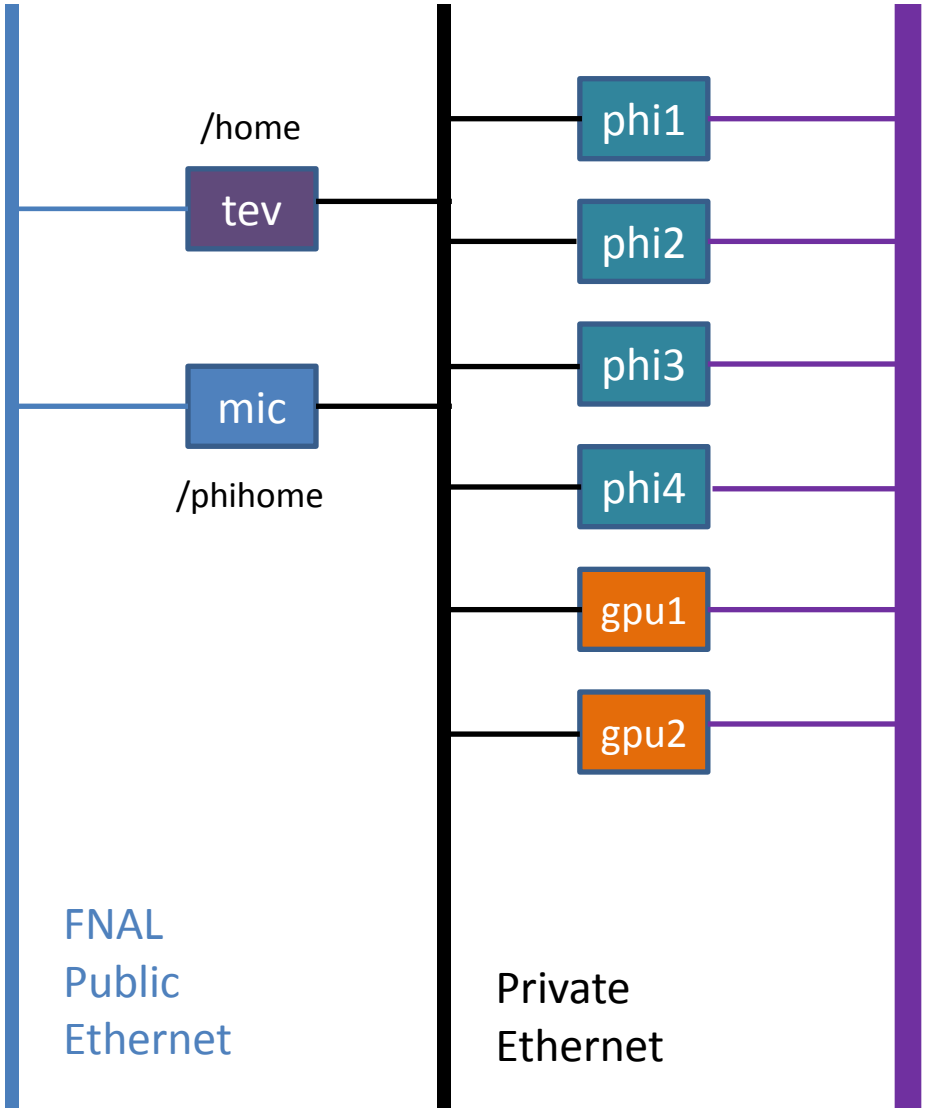
Xeon Phi Performance

- Each core has a 64-byte-wide vector floating point unit that can retire 16 single-prec. or 32 double-prec. floating point instructions/clock cycle @ 1.05 GHz → 1.01 TFlop/sec peak
  (2.02 TFlop/sec peak single prec.)
  - Currently supported only by Intel compilers
  - Open source LLVM compiler will likely support the MIC instruction set within the next year
- Achieving good performance requires:
  - Vectorized code to use these FP units
  - Parallel code, either multithreaded or MPI-style, to use all 240 hardware threads cooperatively
  - Double-precision matrix multiply ("DGEMM" from the HPL Linpack benchmark) achieves 0.76 TFlop/sec. Expect this 75% of peak to be the best possible performance achievable on real codes.
  - Moving data to/from host over PCIe can be a significant overhead

NVIDIA Kepler Model K20X

- The latest NVIDIA GPU:
  - 1.31 TFlop/sec double precision
  - 3.95 TFlop/sec single precision
  - 6 GB of on-board GDDR5 memory
  - 250 GB/sec peak memory bandwidth
  - 2688 CUDA cores
- Compatible with existing CUDA codes
  - Tuning will be necessary to achieve best performance
- See my Feb 20, 2012 AEM special report for a discussion of GPUs, CUDA, and applications of interest to Fermilab

Prototype Phi/Kepler Cluster



- Prototype cluster nodes attached to the existing "Wilson" cluster operated by the HPC Department (login node = tev.fnal.gov)
- 4 Phi servers, each with 4 Xeon Phi cards
- 2 GPU servers, each with a Kepler K20X
- Prototype nodes can be accessed only through the batch system on tev.fnal.gov
- Infiniband supports multi-Phi or multi-Kepler parallel codes
- Software tools (compiler, CUDA) available on mic.fnal.gov
- In FY14, plan to add a Xeon Phi 7120P to each GPU server (61 cores, 16 GB, 1.24 GHz)

Phi and Kepler Applications

- Areas of interest to Fermilab include:
  - Lattice QCD
  - Accelerator simulations
  - GEANT4
  - Multi-core software frameworks
  - Online triggers, and specifically the Mu2e online filter
  - CMS online and offline processing
  - DES processing (2-point correlation functions)
  - Event generation
- Representatives of most of the above areas have already begun exploring the Phi cards on this cluster.

For More Information

- To request an account, send mail to
  tev-admin@fnal.gov
- Preliminary documentation is available:
  http://tev.fnal.gov/phigpu.shtml
- Mailing list for FNAL developers:
  tev-devel@fnal.gov